

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-157091  
 (43)Date of publication of application : 31.05.2002

(51)Int.CI. G06F 3/06  
 G06F 12/16

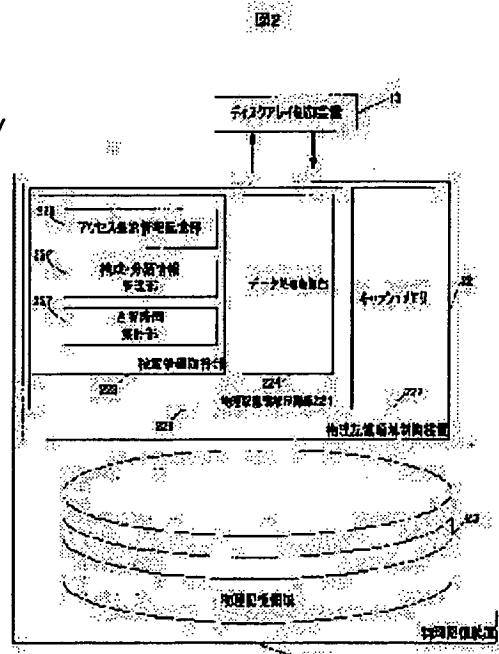
(21)Application number : 2000-353010 (71)Applicant : HITACHI LTD  
 (22)Date of filing : 20.11.2000 (72)Inventor : EGUCHI KENTETSU  
 MOGI KAZUHIKO  
 ARAKAWA TAKASHI  
 OEDA TAKASHI  
 ARAI HIROHARU

## (54) STORAGE SUB-SYSTEM, AND MEMORY USED THEREFOR

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To obtain an occupied time of a logic storage area in a physical memory, and to obtain precise access occupied time information in every I/O to the physical memory.

**SOLUTION:** A physical storage area controller 22 on the individual physical memory 15 is provided with a table 225 for storing information about access requirement from a host computer, a table 227 for totaling the occupied time as to access, a table 226 for control information for classifying constitution of a disk array, and a data processing control part 224 for obtaining constitution information and classification information of the logic storage area form a disk array controller 13, and for requesting the constitution information and the classification information of the logic storage area to the disk array controller, when necessary. The disk array controller 13 is provided with a means for transmitting the constitution information of the disk array at the present time to the physical storage area controller in response to the request from the physical storage area controller on the physical memory.



### LEGAL STATUS

[Date of request for examination] 18.07.2003

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision  
of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office





【0026】物理記憶部2.1は、主に線形増  
幅器2.2から読み込む。これにより、物理記憶部1.5の処理  
能性を上げることができる。

【0029】図3はストレージアレイ制御装置の処理動作を説明する。図3のディスクアブシステムが起動されると、  
まず、時間情報を「1／0処理の処理別毎（ランダムアクセスの時間情報とスカラーシェーベンシャルアクセスの時間情報）」に分類して占有時間  
合算部22に記録する。次に、1／0処理によって、物理記憶領域にどのくらいの時間アセスし  
たか、あるいは、物理記憶装置15のキャッシュメモリにどのくらいの時間の集計を行なうことが可能となる。

[0030] (1) ストレージサブシステム 1 のデータ処理制御部 2.2.4 は、ディスクアレイ制御接続装置 1.3 にデータを送信する。また、データ処理制御部 2.2.4 は、受信した命令がライ特権命令であった場合は、データ処理制御部 2.2.4 は、データをキャッシュメモリ 2.2.2 に保存すれば、キャッシュメモリ 2.2.2 にデータを読み出し、キャッシュメモリ 2.2.2 にデータを読み出せば、物理記憶領域 2.3 にデータを読み出せる。データ処理制御部 2.2.4 は、ディスクアレイ制御接続装置 1.3 から送信されるデータのリードまたはライ特権命令では、データをキャッシュメモリ 2.2.2 に保存すれば、キャッシュメモリ 2.2.2 にデータを読み出せば、物理記憶領域 2.3 にデータを読み出せる。

[0031] (2) 次に、ディスクアレイ制御接続装置 1.5 は、前述した情報の送信により、物理記憶装置 1.5 とセス可能となつたときに、物理記憶装置接続装置 1.5 がアクセス可能な場合から送られるべき物理記憶接続装置 1.5 が接続する。このとき、物理記憶接続装置 1.5 は、ディスクアレイ接続情報 1.4 を送信する。このとき、物理記憶接続装置 1.5 は、データを読み出したときに、物理記憶接続装置 2.3 に書き込みます。直物記憶接続装置 2.3 に書き込みます。

もよい。	〔0032〕(3) 続いて、ディスクアレイ制御装置	3は、1／0バス 18経由でホスト1よりストレ
〔0032〕(4) 前記受信データとしてホスト1	サブシステム1 12に、そのストレージサブシステム	内の論理記憶領域に対応するドライブや处理器
〔0033〕(1) 前記受信データとしてホスト1	を受信した場合、ディスクアレイ制御装置1 3は、	1／0バスを通じて、データを受け渡すもの等の様々なデータを受信する(ステップ3 30)。
〔0033〕(2) 前記受信データとしてホスト1	を受信した場合は、論理記憶領域に対するリ	ト1／0により指定された論理記憶領域に対するリ
〔0033〕(3) 前記受信データとしてホスト1	を受信した場合は、ドライブアドレス(物理ア	ドレス)を用いて物理記憶領域のアドレス(物
〔0033〕(4) 前記受信データとしてホスト1	を受信した場合は、物理記憶領域に対するア	ドレス)に変換する論理／物理対応情報を1 4 1を用
〔0033〕(5) 前記受信データとしてホスト1	を受信した場合は、論理記憶領域アドレスに対応する物理記憶領域ア	ドレスを用いて物理記憶領域アドレスに対応する物理記憶領域アドレスを求める(ステップ3 40、3 50)
〔0034〕(1) ディスクアレイ制御装置1 3は	、その論理記憶領域アドレスを用いて物理記憶領域ア	ドレスを用いて物理記憶領域アドレスを求める(ステップ3 40、3 50)
〔0034〕(2) ディスクアレイ制御装置1 3は	、論理記憶領域アドレスを用いて物理記憶領域ア	ドレスを用いて物理記憶領域アドレスを求める(ステップ3 40、3 50)
〔0034〕(3) ディスクアレイ制御装置1 3は	、論理記憶領域アドレスを用いて物理記憶領域ア	ドレスを用いて物理記憶領域アドレスを求める(ステップ3 40、3 50)
〔0034〕(4) ディスクアレイ制御装置1 3は	、論理記憶領域アドレスを用いて物理記憶領域ア	ドレスを用いて物理記憶領域アドレスを求める(ステップ3 40、3 50)
〔0034〕(5) ディスクアレイ制御装置1 3は	、論理記憶領域アドレスを用いて物理記憶領域ア	ドレスを用いて物理記憶領域アドレスを求める(ステップ3 40、3 50)

リード処理の場合、前述の物理アドレスを有するリードデータを読み出し、ホスト10物理記憶接続からリードデータを読み出し、ホスト10データを転送し、ライト処理の場合、ホスト1データを転送されたライトデータを受信し、その物理アドレスを有することにより、物理記憶接続15内で1/O処理による論理記憶接続、物理記憶接続23、あるいは、キャッシュメモリ22にアクセスしたときのアクセス時

【0041】図5はディスクアレイ制御装置1、記憶装置1、5の線形情報取扱部2、2、3内の情報処理装置1、5の処理動作を説明するフローチャートであり、以下、これについて説明する。

【0036】(1) ディスクアレイ制御装置1は、物理記憶装置1、5の構成やRAIDレベルの変化、論理記憶領域が現在ある物理記憶領域アドレスとは別物理記憶領域のアドレスに移動する等によって、論理記憶領域のアドレスと物理記憶領域のアドレスの対応が変化したことを監視し、変化があった場合、再度、物理記憶装置

【0042】(1) ディスクアレイ制御装置1、トレーシングシステム12が起動された後、記憶装置1、5に有時間情報1、4を初期化し、その後、接続さるフローチャートであり、以下、これについて

る物理記憶装置1にその物理記憶装置1

リード／ライト処理部 330で用いる論理記憶領域を示すアドレスである。また、物理アドレスは、実際データが格納される物理記憶装置 15上の傾斜を示すアドレスであり、物理記憶装置番号及び物理記憶装置内アドレスからなる。記憶装置番号は、個々の物理記憶装置 15を示す。記憶装置番号は、物理記憶装置 15内で記憶領域を示すアドレスである。



り、より最適なストレージサブシステムの性能のチューニングを行うことができる。

## 【図1】

【図1】本発明によるストレージサブシステムを示すブロック図である。  
計算機システムの構成を示すブロック図である。

## 【図2】

物理記憶装置の構成を示すブロック図である。

## 【図3】

ストレージサブシステムが起動されたときのディスクアレイ制御装置の処理動作を説明するフロー図である。

## 【図4】

物理記憶領域のアドレスと物理記憶領域のアドレスに対する動作を説明するフローチャートである。

## 【図5】

ディスクアレイ制御装置が物理記憶装置の取扱い装置動作を実行するフローチャートである。

## 【図6】

物理記憶装置内の物理記憶装置制御装置の処理動作を説明するフローチャートである。

## 【図7】

ディスクアレイ制御装置内に保持されている論理記憶領域と物理記憶領域とのアドレスの対応を管理する動作を説明するフローチャートである。

## 【図8】

物理記憶装置の構成例と物理記憶装置使用状況情報を記述する範囲を示す図である。

## 【図9】

物理記憶領域内の構成・分割情報管理部における構成要素とのアドレスの対応を示す図である。

## 【図10】

物理記憶装置内の取扱情報取扱部の占有時間割付部に保管・格納される記憶領域へのアクセスによる占有時間情報のデータベースの構成例を示す図である。

## 【図11】

占有時間情報のデータベースの構成例を示す図である。

## 【図12】

占有時間集計部の構成例を示す図である。

## 【図13】

物理記憶装置の構成例を示す図である。

## 【図14】

物理記憶装置の構成例を示す図である。

## 【図15】

物理記憶装置の構成例を示す図である。

## 【図16】

物理記憶装置の構成例を示す図である。

## 【図17】

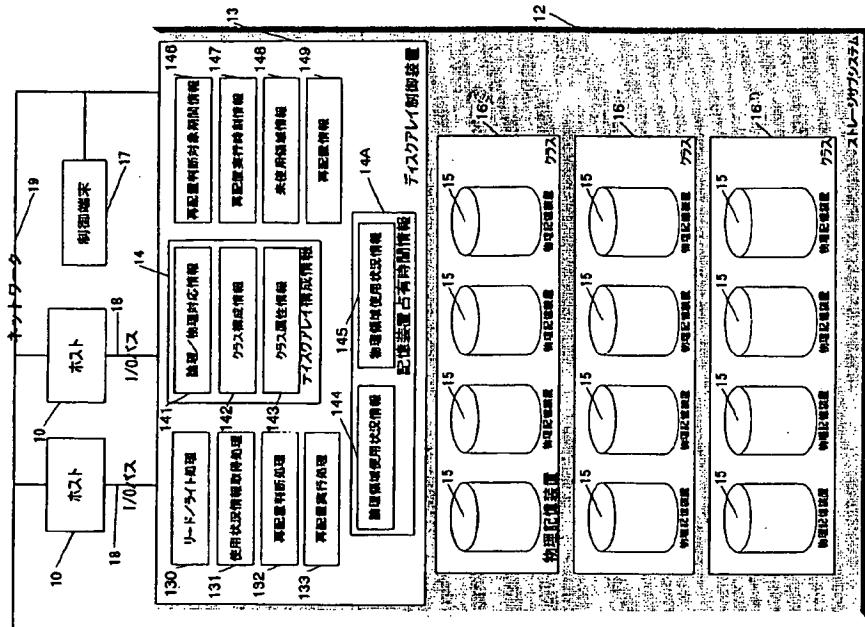
物理記憶装置の構成例を示す図である。

## 【図18】

物理記憶装置の構成例を示す図である。

## 【図19】

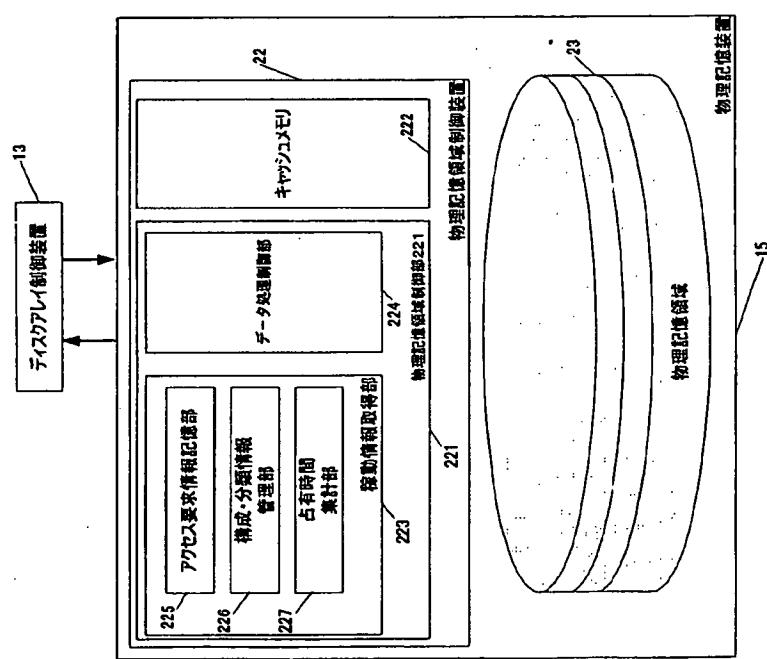
物理記憶装置の構成例を示す図である。



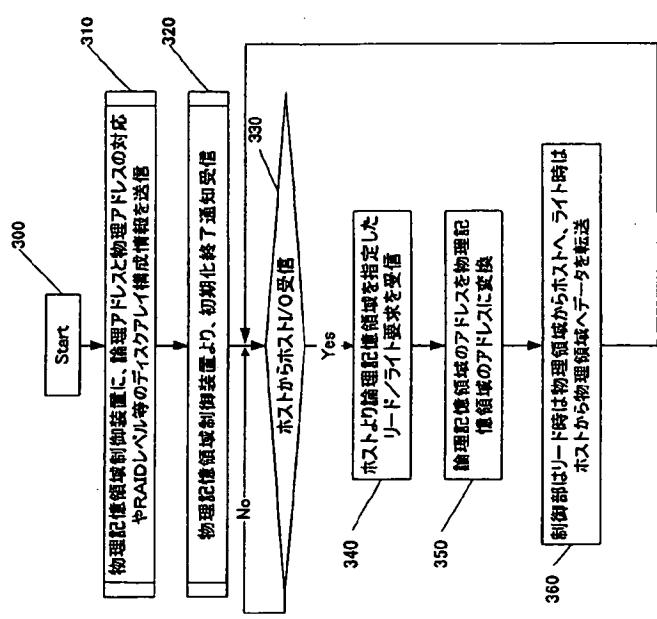
【図8】

記憶領域 番号	I/O ジョブ 番号	Sequential Read	Sequential Write Priority	Random Read	Random Write Priority	Random Write Data	キャッシュ ヒット	合計	
								占有時間	占有時間
1								占有時間	占有時間
2								占有時間	占有時間
3								占有時間	占有時間
4								占有時間	占有時間
...								...	...

【図2】  
図2



【図3】  
図3

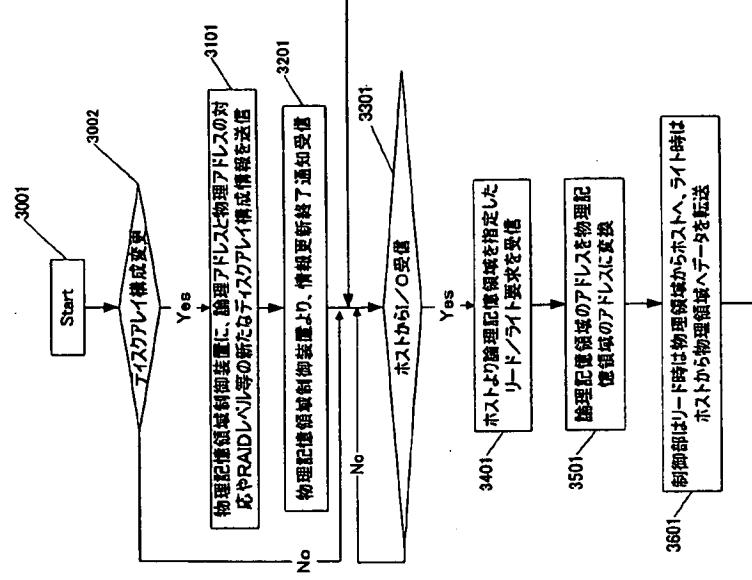


【図9】  
図9

論理記憶領域番号	論理アドレス	物理記憶装置番号	物理記憶領域アドレス
1	3000~3999	1	0~999
4	4000~4999	1	1000~1999
5	5000~5999	1	2000~2999

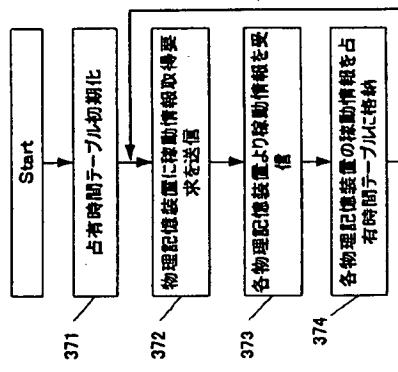
【図4】

図4



【図5】

図5



【図6】

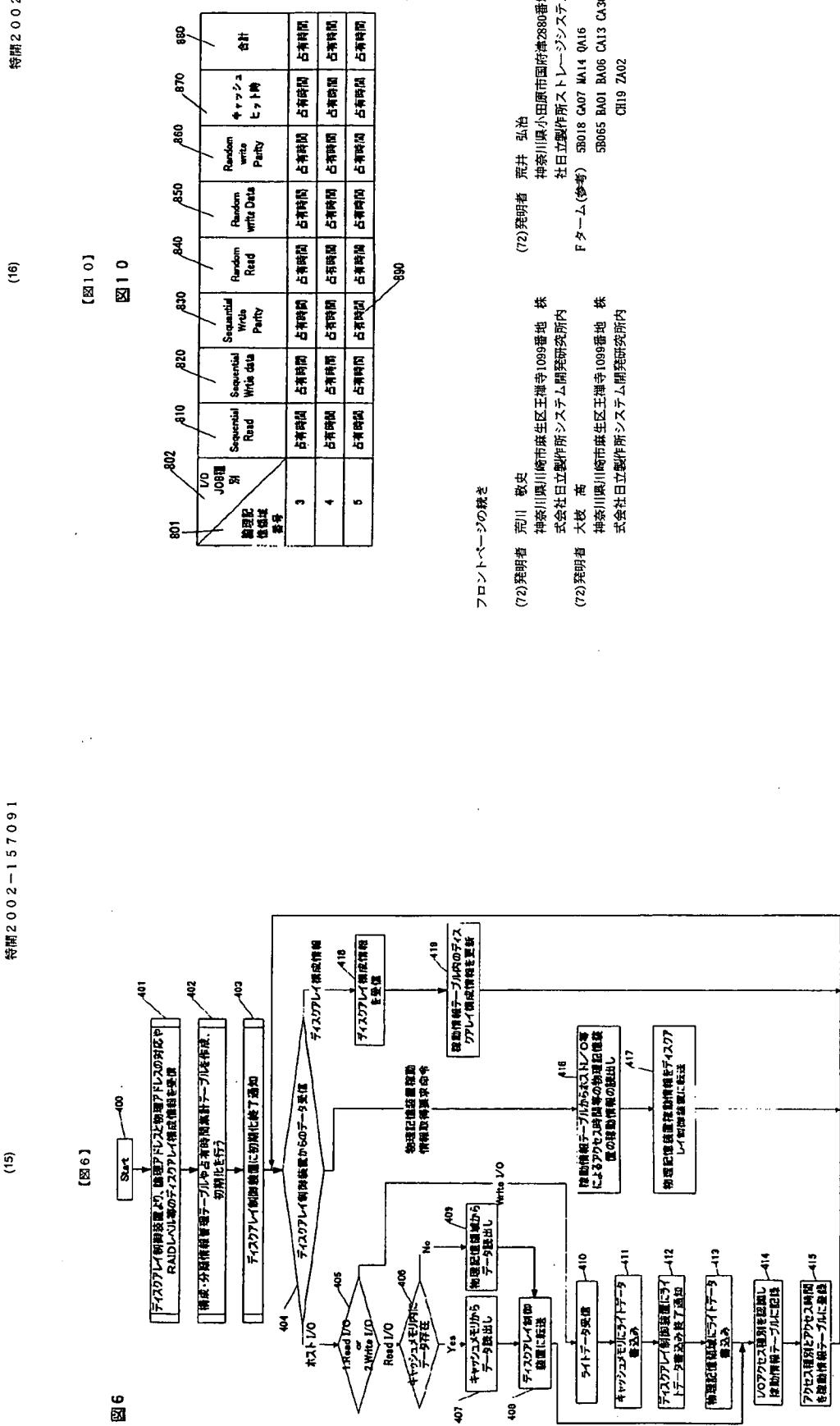
図6

物理記憶装置番号	物理記憶装置アドレス	物理記憶装置番号	物理記憶装置アドレス	レイドレベル	バーティカル番号
0	0~999	0	0~999	1	100
1	1000~1999	0	1000~1999	1	100
2	2000~2999	0	2000~2999	1	100
3	3000~3999	1	0~999	5	120
4	4000~4999	1	1000~1999	5	120
5	5000~5999	1	2000~2999	5	120
...	...	...	...	...	...

【図7】

図7

物理記憶装置番号	物理記憶装置アドレス	物理記憶装置番号	物理記憶装置アドレス	レイドレベル	バーティカル番号
0	0~999	0	0~999	1	100
1	1000~1999	0	1000~1999	1	100
2	2000~2999	0	2000~2999	1	100
3	3000~3999	1	0~999	5	120
4	4000~4999	1	1000~1999	5	120
5	5000~5999	1	2000~2999	5	120
...	...	...	...	...	...



Copied from 10633480 on 03/15/2006

## \* NOTICES \*

JPO and NCIP are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.  
2.\*\*\*\* shows the word which can not be translated.  
3.In the drawings, any words are not translated.

## CLAIMS

## Claim(s)

[Claim 1] A means to be connected to 1 or two or more computers, and to acquire the operating condition information on two or more physical memory equipments and two or more of these physical memory equipments. In the storage subsystem which has a means to perform matching with the logic storage region which said computer makes a read/write object, and the physical memory field of said physical memory equipment It is the storage subsystem which each of two or more of said physical memory equipments is equipped with a physical memory field control unit, and is characterized by this physical memory field control unit having a means to acquire the operating condition of a physical memory field.

[Claim 2] A means to be connected to 1 or two or more computers, and to acquire the operating condition information on two or more physical memory equipments and two or more of these physical memory equipments. In the storage subsystem which has a means to perform matching with the logic storage region which said computer makes a read/write object, and the physical memory field of said physical memory equipment A means to acquire the operating condition information on said two or more physical memory equipments, and a means to perform matching with the logic storage region which said computer makes a read/write object, and the physical memory field of said physical memory equipment It is prepared in the control unit which controls one or more above-mentioned physical memory equipments. Said control unit Furthermore, it has a means to transmit the information which performed matching with the logic storage region of physical memory equipment, and the physical memory field of physical memory equipment to said two or more physical memory equipments of each. It is the storage subsystem which each of two or more of said physical memory equipments is equipped with a physical memory field control unit, and is characterized by this physical memory field control unit having a means to acquire the operating condition of a physical memory field.

[Claim 3] It is physical memory equipment which is equipped with a physical memory field control unit in the physical memory equipment which constitutes a storage subsystem, and is used for the storage subsystem according to claim 1 or 2 characterized by this physical memory field control unit having a means to acquire the operating condition of a physical memory field.

[Claim 4] Said physical memory field control unit is physical memory equipment according to claim 3 characterized by having further a means to store the operating condition information on the acquired physical memory field.

[Claim 5] Said physical memory field control unit is physical memory equipment according to claim 3 or 4 characterized by to have further a means store the information which matched the logic storage region and the physical memory field of the physical-memory equipment which receives the operating condition information on the physical memory field of self-physical memory equipment from a means to transmit to said control unit, and said control unit, according to the acquisition demand of the operating-condition information on a physical-memory field which receives from said control unit.

[Translation done.]

## \* NOTICES \*

JPO and NCIPI are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.\*\*\*\* shows the word which can not be translated.

3.In the drawings, any words are not translated.

## DETAILED DESCRIPTION

## [Detailed Description of the Invention]

## [001] [Field of the Invention]

This invention relates to the store used for a storage subsystem and its system, and relates to the store used for the storage subsystem which has two or more stores especially, and its system.

## [002] [Description of the Prior Art]

The disk array system is known as a highly efficient secondary storage system used for a computer system.

[003] A disk array system is a system which made it possible to perform read/write of the data which arrange two or more physical memory equipments in the shape of an array, divide and store data in each physical memory equipment, and said each physical memory equipment is operated to juxtaposition, and are stored in said each physical memory equipment by dividing at a high speed.

[004] As a conventional technique about a disk array system, the technique indicated by DCG Patterson, G.Gibson, and RH Kats, "A Case for Redundant Arrays of Inexpensive Disks (disk array)" etc. (inch Proc. ACM SIGMOD, pp.109-116, and June 1988, etc. is known. This conventional technique gives the classification of level 5 from level 1 to the disk array system which added redundancy according to that configuration. Moreover, a disk array system without redundancy may be added to these classification, and this may be called level 0 to them. In order to realize each above-mentioned level as a different configuration according to redundancy etc., cost differs from performance characteristics etc. And in building a disk array system, the array (group of physical memory equipment) of two or more level is made intermingled in many cases. Here, the group of the disk array which added redundancy is called a parity group. Moreover, in order to realize optimal cost performance in cost's changing with the engine performance, capacity, etc. also about physical memory equipment, and building a disk array system, two or more sorts of physical memory equipments with which the engine performance differs from capacity too may be used.

[005] It distributes to said above physical memory equipments, and the data stored in a disk array system are arranged. For this reason, a disk array system needs to perform matching of the physical memory field which shows the logic storage region which the host computer connected to a disk array system accesses, and the storage region of said physical memory equipment, i.e., address translation.

[006] The technique indicated by JP-9-274544,A etc. is known as a conventional technique about the disk array system which processes address translation. This conventional technique says that a means to acquire the information about I/O access over the logic storage region from a host computer, and a means to change matching with the physical memory field of a logic storage region, and to perform physical relocation realize the optimal arrangement of the stored data, and the technique in which a disk array control unit acquires the I/O access occupancy hour entry to a logic storage region is indicated by this official report.

[007] In order to perform the load distribution of a disk array system, in case it changes matching with the physical memory field of a logic storage region and performs physical matching

relocation, since the I/O access occupancy hour entry to a logic storage region serves as data which become origin, it is important.

## [008]

[Problem(s) to be Solved by the Invention] Although, as for the conventional technique indicated by the official report mentioned above, it is said that a disk array control unit acquires the occupancy time amount by I/O to a logic storage region, the approach shown in this conventional technique has the trouble that it explains below.

## [009]

First, the case where the writing (light) of data is performed to a certain logic storage region is considered. In this case, the light of data is performed to the physical memory field corresponding to the logic storage region concerned. A physical memory field is in physical memory equipment, and physical memory equipment is constituted by the cache memory and the physical memory field which mainly carry out the cache of a physical memory field control section and the data. And when carrying out the light of the data to a physical memory field, it writes in, when a physical memory field control section writes light data in a cache, and the response of termination is notified to a disk array control unit. For this reason, in order that the conventional technique mentioned above may actually carry out the light of the data, it will have the trouble that the time amount which accessed the physical memory field is not known.

## [010]

Next, the case where reading (lead) of data is performed to a certain logic storage region is considered. In this case, in fact, although the lead of the data in the physical memory field corresponding to the logic storage region concerned is performed, when that lead data is in the cache memory in physical memory equipment, a physical memory field is not accessed, but cache memory is accessed, and that data is returned. For this reason, it will have the trouble that the exact time amount which could not distinguish and accessed the physical memory field does not understand whether the conventional technique mentioned above actually had access in the physical memory field for the data lead.

## [011]

Moreover, the case where the logic storage region currently crossed to A, B, C, D, and multiple times has access is considered. And the response of A' and B is made into B' etc. for the response of A, and time of day of B (t) and response A' is made [the time of day of Access A] into A' (t) etc. for the time of day of A (t) and Access B. Here, Access A accesses physical memory equipment with a data lead, leads data, and assumes Access B to be what had the response B' without accessing physical memory equipment, since data are in a cache with a data lead. In this case, the direction of response B' of Access B which came after response A' of the access A which suited previously becomes previously. That is, it becomes A(t) < B (t), A' (t) > B' (t). At this time, the conventional technique which acquires the occupancy time amount by I/O to a logic storage region in a disk array control unit produces the trouble that sweet red bean soup with moth cannot perform whether which I/O accessed the physical memory field in physical memory equipment how much, or hit into the cache of physical memory equipment.

## [012]

The purpose of this invention is by solving the trouble of the conventional technique mentioned above and acquiring the occupancy time amount of a logic storage region with physical memory equipment to offer the store which offers the storage subsystem which enabled it to acquire the occupancy time amount (real operating time) of each logic storage region by the system configuration unacquirable [with a disk array control unit], and is used for this.

## [013]

The purpose of this invention moreover, is by taking into consideration the effect on the physical memory equipment for every I/O in order to be able to make small the error of the utilization factor of the logic storage region of said disk array system, or utilization factor prediction and to perform optimum performance tuning more. It is in offering the store which offers the storage subsystem which can acquire the access occupancy hour entry for every I/O to physical memory equipment with a more high precision, and is used for this.

## [014]

[Means for Solving the Problem] According to this invention, said purpose is connected to 1 or two or more computers. Two or more physical memory equipments, In the storage subsystem which has a means to acquire the operating condition information on two or more of these physical memory equipments, and a means to perform matching with the logic storage region which said computer makes a read/write object, and the physical memory field of said physical

**memory equipment A** means to acquire the operating condition information on said two or more physical memory equipments, and a means to perform matching with the logic storage region which said computer makes a read/write object, and the physical memory field of said physical memory equipment it is prepared in the control unit which controls two or more above-mentioned physical memory equipments. Said control unit Furthermore, it has a means to transmit the information which performed matching with the logic storage region of physical memory equipment, and the physical memory field of physical memory equipment to said two or more physical memory equipments of each. It has a physical memory field control unit, and this physical memory field control unit is attained by having a means of two or more of said physical memory equipments to acquire the operating condition of a physical memory field, respectively. [0015] Moreover, said purpose is set to the physical memory equipment which constitutes a storage subsystem. A means by which have a physical memory field control unit and this physical memory field control unit acquires the operating condition of a physical memory field. An acquisition demand of the operating condition information on a means to store the operating condition information on the acquired physical memory field, and the physical memory field received from said control unit is accepted. It is attained by having a means to transmit the operating condition information on the physical memory field of self-physical memory equipment and a control unit, and a means to store the information which matched the logic storage region and a physical memory field of the physical memory equipment received from said control unit. [0016]

[Embodiment of the Invention] Hereafter, a drawing explains the operation gestalt of the storage used for the storage subsystem by this invention, and its system to a detail.

[0017] The block diagram showing the configuration of the computer system with which drawing 1 is equipped with the storage subsystem by this invention, and drawing 2 are the block diagrams showing the configuration of physical memory equipment. In drawing 1 R>1 and drawing 2, 1 host and 12/10 A storage subsystem, 13 disk array control information and 15 for a disk array control device and 14 Physical memory equipment. In 16, a disk array and 17 an I/O bus are 19 for a control terminal and 23 The read/write processing section, 131 the relocation decision memory field control unit and 23 The read/write processing section, 131 the relocation decision processing section and 133 for the operating condition information acquisition processing section and 132 The relocation executive operation section, 141 class configuration information and 143 to the information corresponding to logic/physics, and 142 Class attribute information. 144 physical field operating condition information and 146 for logic field operating condition information and 145 Relocation decision horizon information, 147 free-space information and 149 for relocation activation time information and 148 Relocation information, 14A — a storage occupancy hour entry and 221 — for the operation information storage section and 224, as to 225 are [ a physical memory field control section and 222 / cache memory and 223 / configuration / classification Research and Data Processing Department and 227 ] the occupancy time amount total sections.

[0018] The computer system shown in drawing 1 consists of 1 which is the calculating machine of a high order or two or more hosts 10, a storage subsystem 12, and a control terminal 17. It connects with the storage subsystem 12 by I/O bus 18, and a host 10 publishes lead of data, and I/O for light processing to the storage subsystem 12. In case this I/O is performed, a host 10 specifies the logical storage region of the storage subsystem 12. That is, a host 10 accesses with the address of a usually logical storage region to the data in a storage subsystem. Moreover, I/O bus 18 is constituted by ESCON, SCSI, the fiber channel, etc.

[0019] The storage subsystem 12 consists of a disk array control unit 13 and two or more physical memory equipments 15. The disk array control unit 13 is equipped with the read/write processing section 130, the operating condition information acquisition processing section 131, the relocation decision processing section 132, and the relocation executive operation section 133, and these processing sections process read/write processing, operating condition information acquisition processing, relocation decision processing, relocation executive operation, etc. Moreover, the disk array control non-dense of the storage subsystem 12 holds the

information 141 corresponding to a logic storage region / physical memory field, the class configuration information 142, the disk array configuration information 1400 of class attribute information 143 grade, storage occupancy hour entry 14A of the logic field operating condition information 144 and physical field operating condition information 145 grade, the relocation decision horizon information 146, the relocation activation time information 147, the free-space information 148, and relocation information 149 grade. In addition, others and parity group information, RAID level information, etc. which were mentioned above may be included in the disk array configuration information 14 mentioned above. [ information ]

[0020] Moreover, the host 10, the disk array control unit 13, and the control terminal 17 are mutually connected by the network 19. A network 19 may be constituted by Ethernet (trademark) FDDI, the fiber channel, etc. A control terminal 17 is usually used in order to perform maintenance, management, etc. of the storage subsystem 12.

[0021] Moreover, in explanation of the operation gestalt of this invention, although it exists, respectively, since it is not important, the component which surely exists in computers, such as memory for performing processing which appears in a host 10, the disk array control unit 13, and a control terminal 17, respectively, and CPU, is not specified here.

[0022] The class division of two or more physical memory equipments 15 formed in the above-mentioned storage subsystem 12 is carried out for every engine performance of physical memory equipment, and they constitute the disk array 16 for every class. Moreover, although not shown clearly, two or more physical memory equipments are used, and the parity group is constituted here. And each of physical memory equipment 15 is constituted by the physical memory field control device 22 which controls the physical memory field 23 and this physical memory field 23 to be shown in drawing 2, and various data are stored in the physical memory field 23. [0023] Moreover, if the address of the physical memory field 23 seems to have mentioned above directly from a host 10, it does not break, but a host 10 accesses the data on two or more logical storage regions on two or more physical memory fields 23. Namely, a host 10 accesses by specifying a logic storage region to the data in the storage region of each physical memory equipment 15 in the storage subsystem 12.

[0024] The disk array control unit 13 is connected with two or more physical memory equipments 15. The lead and light processing instruction I/O which controlled two or more physical memory equipments 15, or were emitted by said host 10 Make the address of the logic storage region where the appointed data exist, and the address of a physical memory field with the address of the logic storage region correspond, transmit data I/O to suitable physical memory equipment 15, and if it is light processing. The data transmitted by the host 10 are transmitted to physical memory equipment 15, and if it is lead processing, the data transmitted from physical memory equipment 15 are received, and it is processing transmitting to a host 10 etc.

[0025] The physical memory field control unit 22 which it has in physical memory equipment 15 is constituted by the physical memory field control section 221 and cache memory 222. Cache memory 222 has the quick rate of processing of the read/write of data compared with the physical memory field 23. And cache memory 222 is used as follows about the data about the lead or light instruction transmitted from the disk array control device 13. That is, in case the light data transmitted from the disk array control device 13 are written in the physical memory field 23 in light processing, data are written also in cache memory 222. Moreover, in case data reading appearance is carried out from the physical memory field 23 in lead processing, when it was written in cache memory 222, or the same data are in cache memory and it comes to physical memory equipment from the disk array control device 13 as a lead instruction to the data by former lead processing, the read data do not read the data from the physical memory field 23, but read it from cache memory 222. Thereby, the processing engine performance of physical memory equipment 15 can be improved.

[0026] The physical memory field control section 221 is mainly equipped with the operation information acquisition section 223 and the data-processing control section 224, and is constituted. The data-processing control section 224 receives the lead or light instruction of data transmitted from the disk array control device 13. And the data-processing control section

224 will read the lead data from cache memory 222, if cache memory 222 is accessed and the lead data exists in cache memory 222, when the received instruction is a lead instruction, if the data is not in cache memory 222, it will access the physical memory field 23, will read the lead data, and will transmit data to the disk array control device 13. Moreover, the data-processing control section 224 is after that, and writes the data in the physical memory field 23 at the same time it writes the data transmitted from the disk array control device 13 in cache memory 222, when the received instruction is a light processing instruction. You may not write light data in cache memory 222, but may also write them in the direct physical memory field 23.

[0027] The operation information acquisition section 223 is constituted by the access request information storage section 225, configuration / classification Research and Data Processing Department 226, and occupancy time amount total section 227 grade. When a logic storage region with the data specified by the I/O process, the physical memory field 23 where the data exists, or cache memory 222 is accessed, the above-mentioned data-processing control section 224 classifies the hour entry of the access for every (random access, sequential access, etc.) processing classification of an I/O process, and records it on the occupancy time amount total section 227. Moreover, from the disk array control device 13, the data-processing control section 224 receives information, such as correspondence information on the address of the logic storage region in a disk array, and the address of the physical memory field 23, and engine performance of physical memory equipment 15, and records it on configuration / classification Research and Data Processing Department 226. Furthermore, the data-processing control section 224 receives a lead of data or light instruction data transmitted from disk array control device 13 grade, and records the instruction data on the access request information storage section 225 so that it may be possible to receive two or more I/O processes.

[0028] The logic storage region according to an I/O process within physical memory equipment 15 when physical memory equipment 15 has a configuration which was mentioned above, the physical memory field 23. Or it becomes possible to classify the hour entry at the time of access when accessing cache memory 222 for every (random access, sequential access, etc.) processing classification of an I/O process, and to record on the occupancy time amount total section 227. It classifies into a physical memory field to what whether time amount access was carried out and physical memory equipment 15 of cache memory 22, hit according to an I/O process, and it becomes possible to total the occupancy time amount.

[0029] Drawing 3 is a flow chart explaining processing actuation of a disk array control device when a storage subsystem is started, and explains this hereafter.

[0030] (1) The disk array control unit 13 transmits the information 141 corresponding to the logic / physics which is the matching information on the address of the logic storage region in the physical memory field 23 in physical memory equipment 15, and the address of the physical memory field where the logic storage region actually exists, the class configuration information 142 and the disk array configuration information 14 of class attribute information 143 grade to the physical memory equipment 15 connected with self-equipment 13 at the time of starting of the storage subsystem 12 (steps 300 and 310).

[0031] (2) Next, the disk array control unit 13 receives the notice the physical memory equipment 15 sent from the physical memory field control unit 22 changed [the notice] to the accessible ready state, when physical memory equipment 15 becomes accessible by transmission of the information mentioned above. Things come and physical memory equipment 15 is in the condition of the initialization termination by the disk array configuration information 141 corresponding to the disk array control device 13 at the time of the initialization termination by the disk array control device 13 (step 320).

[0032] (3) Then, the disk array control device 13 receives data with more various what has transmitted host I/O of a lead or light processing to the storage subsystem 12 to the logic storage region in the storage subsystem 12, things which deliver an instruction and data with disk array control devices than a host 10 by I/O-bus 18 course (step 330).

[0033] (4) When host I/O is received as said received data, the disk array control unit 13 receives the lead or light demand to the logic storage region specified by host I/O, and asks for the logic storage region address and the corresponding address of the physical memory field 23 using the information 141 corresponding to the logic / physics which changes the address (logical

address) of the logic storage region into the address (physical address) of a physical memory field (steps 340 and 350).

[0034] (5) The disk array control device 13 specifies the address of the physical-memory field where predetermined data exist, and, in lead processing, transmits light data to the physical-memory equipment which transmits lead data for the physical memory equipment which has the above-mentioned physical address to lead data to read-out and a host 10, receives the light data transmitted by the host 10 in light processing, and has the physical address (360).

[0035] Drawing 4 is a flow chart explaining processing actuation of a disk array control device when correspondence of the address of a logic storage region and the address of a physical memory field changes, and explains this hereafter.

[0036] (1) By the change in physical memory equipment 15, change of RAID level, and a logic storage region moving the disk array control device 13 to the address of a physical memory field different from a certain physical memory \*\*\* address now etc. When it supervises that correspondence of the address of a logic storage region and the address of a physical memory field changed and is changeable, physical memory equipment 15 is received again. The information 141 corresponding to the logic / physics which is the matching information on the address of the logic storage region in the physical memory field 23 in physical memory equipment 15, and the address of the physical memory field where the logic storage region actually exists, The class configuration information 142 and the disk array configuration information 14 of class attribute information 143 grade are transmitted (step 310).

[0037] (2) Next, the disk array control unit 13 receives the notice the physical memory equipment 15 sent from the physical memory field control unit 22 changed [the notice] to the accessible ready state, when physical memory equipment 15 becomes accessible by transmission of the information mentioned above. Things come and physical memory equipment 15 is in the condition of the updating termination by the disk array configuration information 14 (step 320).

[0038] (3) Subsequent processing is performed like the case of steps 330, 3400, 3500, and 3600 explained by drawing 3 (steps 3301, 3401, 3501, and 3601).

[0039] In addition, in steps 320 and 3201 mentioned above, as for the disk array control unit 13, \*\* does not need to receive the information that the physical memory field control unit 22 to physical memory equipment 15 changed in the accessible condition. In this case, what is necessary is just to perform access processing for performing predetermined processing to physical memory equipment 15, when it assumes that it is in accessible conditions, such as a lead and a light, to physical memory equipment 15 and the access directions to a certain storage region come for the disk array control unit 13 after a certain fixed time amount. Moreover, when there is no response into waiting and fixed time amount until it performs access processing in order to perform predetermined processing again when there is no response to predetermined access processing, or a response comes back, it is good also as a method which tells that to the module which issued the access directions to a certain storage region.

[0040] Moreover, the information 141 corresponding to logic / physics in the above-mentioned is information to which a logic storage region and a physical memory field are made to correspond. And the logical address is the address which shows the logic storage region which a host 10 uses in said read / write processing section 130. Moreover, a physical address is the address which consists of the physical memory device number and the address in physical memory equipment. A storage number shows each physical memory equipment 15. The address in storage is the address which shows the storage region within physical memory equipment 15. [0041] Drawing 5 is a flow chart explaining processing actuation of the disk array control device 13 at the time of the disk array control device 13 reading the information in the operation information acquisition section 223 of physical memory equipment 15, and explains this hereafter.

[0042] (1) After the storage subsystem 12 is started, the disk array control device 13 initializes storage occupancy hour entry 14A, and transmits an acquisition demand of the access using the information 141 corresponding to the logic / physics which changes the address (logical

equipments 15 connected after that (steps 371 and 372).  
 [0043] (2) Next, the disk array control unit 13 stores the access occupancy hour entry of reception and each physical memory equipment in storage occupancy hour entry 14A for an access occupancy hour entry from each physical memory equipment 15 (steps 373 and 374).

[0044] In addition, the timing of acquisition of the access occupancy information on the disk array control unit 13 mentioned above The method which reads the access occupancy information by access to physical memory equipment 23 from the module of others by host I/O, backup, etc., to a fixed time interval from the occupancy time amount total section 227 in each physical memory equipment 15. When an access occupancy hour entry acquisition demand is transmitted to the disk array control unit 13 from other modules (for example, a host 10 and a control terminal 17), it is various and dependent on a design.

[0045] The access occupancy hour entry acquired by the above-mentioned is recorded on the occupancy time amount total table in the disk array control device 13.

[0046] Drawing 6 is a flow chart explaining processing actuation of the physical memory field control device 22 in physical memory equipment 15, next explains this.

[0047] (1) The physical memory field control device 22 receives the disk array configuration information 14 which is information on the logical address and the physical address in the physical memory field 23 of the data 15 transmitted from the disk array control device 13 at the time of starting of the storage subsystem 12, i.e., physical memory equipment, such as correspondence, (steps 400 and 401).

[0048] (2) The physical memory field control unit 22 which received the disk array configuration information 14 performs creation initialization of configuration / classification information 226 in the operation information acquisition section 223, or the occupancy time amount total section 227 based on the information (step 402).

[0049] (3) After initialization processing of the table in step 402 is completed, in order to make the physical memory equipment 15 recognize an accessible thing, notify that initialization processing was completed to the disk array control unit 13. In addition, it is not necessary to transmit the information to which physical disk equipment changed in the accessible condition to the disk array control device 13. When the access directions to a certain storage region come to the physical memory field control unit 22 of physical memory equipment 15 after the fixed time amount, in this case, the physical memory field control unit 22 If it is in a condition accessible to the physical memory field 23 in order to perform predetermined processing Access the physical memory field 23, perform predetermined processing, and if impossible [whether processing is determined in the condition of having stored access information in the access request information storage section 225, and having become accessible to the physical memory field is performed, and ] or you may make it not receive a certain access directions to a storage region and it comes to resemble the physical memory field 23 in the accessible condition, in order to perform predetermined processing (step 403).

[0050] (4) After that, the physical memory field control device 22 waits to transmit host I/O, and a physical memory equipment operation acquisition demand instruction or new disk array configuration information from the disk array control device 13, and receives it (step 404).  
 [0051] (5) At step 404, if host I/O is received from the disk array control unit 13 when the I/O judges lead processing or light processing and it is lead processing, the data processing control section 224 will confirm whether the data which should be read exist in cache memory 22 (steps 405 and 406).

[0052] (6) When the data is read from cache memory 22 when the data exists in cache memory 22, and the data does not exist in cache memory 22 with the check of step 406, read the data from the physical memory field 23, and transmit data to the disk array control unit 13 (steps 407, 709, and 408).  
 [0053] (7) At step 405, when judged with host I/O being light processing, the data-processing control section 224 receives the light data transmitted by the host 10, and writes the light data in cache memory 22 (steps 410 and 411).

[0054] (8) And the data-processing control section 224 stores the above-mentioned light data in time amount total table in the receiving disk array control device 13. The example shown in

the physical memory field 23 while notifying the notice of data write-in termination to the disk array control device 13 (steps 412 and 413).  
 [0055] The data-processing control section 224 to cache memory 22 after processing of step 408, or processing of step 413 (9) In access \*\* Or the JOB classification information on the information on whether the physical memory field 23 was accessed, a random lead, a sequential lead, etc., The access classification of the JOB classification information on the random read/write at the time of writing light data in the physical memory field 23, sequential read/write, etc. is recognized. The occupancy hour entry which accessed cache memory 22 or the physical memory field 23 is stored in the occupancy time amount total section 227 in the operation information acquisition section 223 for every access classification (steps 414 and 415).  
 [0056] (10) If new disk array configuration information is received from the disk array control device 13 at step 404, the data-processing control section 224 will rewrite the information in configuration / classification Research and Data Processing Department 226, the operation information acquisition section 223, corresponding to new disk array configuration information (steps 418 and 419).

[0057] (11) When a physical memory equipment operation information acquisition demand instruction is received from the disk array control device 13 at step 404, the data-processing control section 224 reads the access occupancy hour entry of the physical memory equipment 15 stored in the occupancy time amount total section 227 in the operation information acquisition section 223, and transmits it to the disk array control device 13 (steps 416 and 417).  
 [0058] In addition, it may be made to perform transmission to the disk array control device 13 from the physical memory equipment 15 of the occupancy hour entry in processing of step 417 mentioned above to the disk array control device 13 with a fixed time interval automatically from physical memory equipment 15. In this case, said physical memory equipment operation information acquisition demand instruction is not transmitted to physical memory equipment 15 from the disk array control unit 13.

[0059] Drawing 7 is drawing explaining the example of a configuration of the information 141 corresponding to logic/physics in the table for managing correspondence with the address of a logic storage region and the address of a physical memory equipment which are held in the disk array control unit 13.  
 [0060] The disk array control unit 13 has managed correspondence with the address of the logic storage region in the physical memory field 23 in two or more physical memory equipments 15 connected, and the address of the physical memory field in the logic storage region. Each of the logic storage region number 500 given to a specific logic storage region, the logical address 510, the physical address 520 by the storage number 521 with a physical storage region with the logic storage region and the address 522 of a physical storage region, the RAID level 530 that shows the engine performance of the physical-memory equipment 15, and the parity group number 540 to which the physical-memory equipment 15 belongs is matched, and the information 141 corresponding to the logic/physics used for this is constituted, as shown in drawing 7 . When processing of a lead, a light, etc. specifies the address of a logic storage region from a host 10, a control terminal 17, and other modules (for example, other disk array control units etc.) and it is accessed to the self-disk array control unit 13 by having such information 141 corresponding to logic/physics, the disk array control unit 13 can change the address of a logic field into the address of a physical storage region, and can perform read/write processing of data correctly to physical memory equipment 15.

[0061] Drawing 8 is drawing showing the example of the logic field operating condition information 144 stored in the disk array control unit 13, and the storage occupancy hour entry 141 of physical field operating condition information 145 grade. Such information is constituted as an occupancy time amount total table.

[0062] The disk array control device 13 reads periodically the access occupancy information on the physical memory field 23 of the physical memory equipment 15 by access from the module of others by host I/O, backup, etc., from the occupancy time amount total section 227 in each physical memory equipment 15, and records the access occupancy hour entry on the occupancy time amount total table in the receiving disk array control device 13. The example shown in

**drawing 8** is every logic storage region number 601 and 1/O. Occupancy time amount is totaled every JOB classification 602. 1/O As a JOB classification, although 670 and a total of 680 are shown by the example of illustration at the time of the sequential lead 610, the sequential light data 620, the sequential light parity 630, the random lead 640, the random light parity 660, and a cache hit, it is 1/O of further others. There may be JOB classification.

[0063] The occupancy time amount read from physical memory equipment 15 may be record by the accumulation value of not only the above-mentioned but the access occupancy time amount for every I/O, and a universal time amount value and a time amount value peculiar to a machine. Moreover, in the disk array control unit 13, the access occupancy time amount of each logic storage region or as physical memory field may be edited, and an occupancy hour entry table may newly be created based on the value which found the access occupancy time amount for every physical memory equipment and every parity group.

[0064] When a host 10 and control terminal 17 grade give the operation information acquisition demand of physical memory equipment to the disk array subsystem 12 by the above-mentioned, even if a host 10 and control terminal 17 grade access direct physical memory equipment 15 and don't acquire the access occupancy hour entry of a logic storage region or a physical memory field, they become possible [ acquiring the access occupancy hour entry of the logic storage region and physical memory field from the disk array control unit 13 ].

[0065] Drawing 9 is drawing showing the example of a configuration of the table which manages matching with the address of a logic storage region and the address of a physical memory field which are stored in configuration / classification Research and Data Processing Department 226 in the operation information acquisition section 223 in physical memory equipment 15.

[0066] From the disk array control device 13, when the relation of the address of a logic storage region and the address of a physical-memory field which are produced by the time of starting of the storage subsystem 12, the change in physical-memory equipment, change of RAID level, migration of a logic storage region, etc. changes, the data-processing control section 224 in physical-memory equipment 15 receives the matching information on the logic storage region and physical-memory field, and stores it in a correspondence table as shows it to configuration / classification Research and Data Processing Department 226 at drawing 9. This correspondence table is constituted by physical ADORE 7520 by the logic storage region number 700 given to a specific logic storage region, the logical address 710, and the storage number 721 with a physical storage region with that logic storage region and the address 722 of a physical storage region. Thereby, it can recognize which address logic storage region physical memory equipment 15 has where of the physical memory regional address in self-physical memory equipment 15.

[0067] Drawing 10 is drawing showing the example of a configuration of the table of the occupancy time amount total section 227 of the operation information acquisition section 223 in physical memory equipment 15.

[0068] This table is every logic storage region number 801 and 1/O. Occupancy time amount is recorded every JOB classification 802. 1/O As a JOB classification, in the example of illustration, although 870 and a total of 880 are shown at the time of the sequential lead 810, the sequential light data 820, the sequential light parity 830, the random lead 840, the random light parity 860, and a cache hit, there may be 1/OJOB classification of further others.

[0069] If the data-processing control section 224 in physical memory equipment 15 has access in a store by 1/O from a host etc. it will accumulate the occupancy time amount 890 in the occupancy time amount total section 227 every JOB classification 802 of access about each logic storage region 801 with access. This becomes possible to obtain the relation between the number of the logic storage region in physical memory equipment 15, the occupancy time amount by the access classification to the logic storage region, and the sum total occupancy time amount within a certain time amount within physical memory equipment 15.

[0070] According to the operation gestalt of this invention mentioned above, the occupancy hour entry of the storage for every I/O can be acquired, and acquisition of the occupancy hour entry of the storage by access to a storage region can be realized within physical memory equipment [0071] According to the operation gestalt of this invention, the thing of two or more physical

memory equipments which constitute a storage subsystem for which the occupancy time amount of the logic storage region of physical memory equipment is acquired becomes respectively possible by the above-mentioned, and the thing of each logic storage region to do for occupancy time amount (real operating time) acquisition becomes possible by the system configuration unacquirable [ with a disk array control unit ]. Moreover, in order according to the operation gestalt of this invention mentioned above to be able to make small the error of the analysis of the utilization factor of the logic storage region of said storage subsystem, or utilization factor prediction and to perform optimum performance tuning more by taking into consideration the effect on the physical memory equipment for every 1/O, the access occupancy hour entry for every I/O to physical memory equipment with a more high precision is acquirable.

[0072] [Effect of the Invention] As explained above, according to this invention, physical memory equipment can acquire the occupancy time amount of a logic storage region, and the occupancy time amount (real operating time) of each logic storage region can be acquired by the system configuration unacquirable [ with a disk array control unit ].

[0073] Moreover, according to this invention, since the access occupancy hour entry for every I/O to physical memory equipment is acquirable, it becomes possible to perform the analysis of the utilization factor of the logic storage region of a storage subsystem and the prediction of a utilization factor in consideration of the effect on the physical memory equipment for every I/O with a small error, and the engine performance of the more nearly optimal storage subsystem can be tuned up.

[Translation done.]

## \* NOTICES \*

JP0 and NCPI are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

## DESCRIPTION OF DRAWINGS

## [Brief Description of the Drawings]

[Drawing 1] It is the block diagram showing the configuration of the computer system equipped with the storage subsystem by this invention.

[Drawing 2] It is the block diagram showing the configuration of physical memory equipment.

[Drawing 3] It is a flow chart explaining processing actuation of a disk array control device when a storage subsystem is started.

[Drawing 4] It is a flow chart explaining processing actuation of a disk array control device when correspondence of the address of a logic storage region and the address of a physical memory fresh changes.

[Drawing 5] It is a flow chart explaining processing actuation of the disk array control device at the time of a disk array control device reading the information on operation information acquisition circles of physical memory equipment.

[Drawing 6] It is a flow chart explaining processing actuation of the physical memory field control device in physical memory equipment.

[Drawing 7] It is drawing explaining the example of a configuration of the information corresponding to the logic/physics which manages correspondence of the address of the logic storage region and physical memory field which are held in the disk array control unit.

[Drawing 8] It is drawing showing the example of storage occupancy hour entries, such as logic field operating condition information stored in a disk array control unit, and physical field operating condition information.

[Drawing 9] It is drawing showing the example of a configuration of the table which manages matching of the address of the logic storage region and physical memory field which are stored in configuration / classification Research and Data Processing Department in physical memory equipment.

[Drawing 10] It is drawing showing the example of a configuration of the table of the occupancy hour entry by access to the storage region accumulated and stored at the occupancy time equipment.

[Description of Notations]

10 Host

12 Storage Subsystem

13 Disk Array Control Unit

14 Disk Array Control Information

15 Physical Memory Equipment

16 Disk Array

17 Control Terminal

18 I/O Bus

19 Network

22 Physical Memory Field Control Unit

23 Physical Memory Field

130 Read/write Processing Section